Laver/Air, eds. Structure and Variation in Influenza Virus

STRUCTURAL STUDIES ON THE HAEMAGGLUTININ GLYCOPROTEIN OF INFLUENZA VIRUS

IAN A. WILSON[+], JOHN J. SKEHEL[++] AND DON C. WILEY[+]
[+]Gibbs Laboratory, Dept. Biochemistry and Molecular Biology, Harvard University, 12 Oxford St., Cambridge, MA 02138, USA: [++]Division of Virology, National Institute for Medical Research, Mill Hill, London, NW7 1AA, England.

SUMMARY

X-ray crystallographic investigation of the haemagglutinin glycoprotein of influenza virus has produced a three-dimensional image of the trimeric molecule at 3Å resolution. A single isomorphous derivative, mercury phenyl glyoxal, was used in the phase determination. The non-crystallographic three-fold axis was determined from the solution of the difference Patterson at 3Å resolution. Three-fold averaging of the electron density map indicates the monomer and molecular boundaries. Interpretation of the map awaits improvement of the phases by non-crystallographic symmetry averaging.[1]

INTRODUCTION

The major surface glycoprotein of influenza virus, the haemagglutinin, has been crystallized[2] and its three-dimensional structure is being determined to 3Å resolution by X-ray crystallographic methods. The haemagglutinin is a trimer composed of two disulphide-linked polypeptide chains, HA_1 and HA_2, including 19% by weight of carbohydrate. The protein can be cleaved from the viral membrane by bromelain, leaving a C-terminal polypeptide "tail" (5000 daltons) embedded in the lipid.[3] Preliminary electron density maps have revealed one bromelain released trimer (about 197,000 daltons) per asymmetric unit of the crystal. Amino acid sequence determination of the complete monomer is well advanced[4] and will greatly facilitate the interpretation of the electron density map. The solution of a mercury phenyl glyoxal difference Patterson and the calculation of a preliminary electron density map using single isomorphous phases and non-crystallographic symmetry averaging[1] is presented.

MATERIALS AND METHODS

Native and derivative crystals. The haemagglutinin glycoprotein of the Hong Kong strain of influenza virus (H3N2),[5] when released by bromelain from the membrane,[3] has been crystallized by microdialysis (50 microlitre) against 1.2-1.45M sodium citrate, 0.1% sodium azide, pH 7.5.[2] Large single octahedral crystals (up to $1mm^3$) are formed with tetragonal space group $P4_1$ or $P4_3$, cell

dimensions a = 163.2Å, c = 177.4Å and one molecule per asymmetric unit (see results for discussion of this observation). A mercury phenyl glyoxal derivative was prepared by placing a preformed crystal into a suspension of 10^{-2}M mercury phenyl glyoxal[6] in the mother liquor of 1.45M sodium citrate, 0.1% sodium azide, pH7.5 for at least 15 hours. A second mercury iodide derivative has been prepared by soaking crystals in 1mM mercury iodide in mother liquor for at least 24 hours at room temperature. Data collection of this second derivative is in progress and no measurements from these crystals have been included in the present structure analysis.

DATA COLLECTION

X-ray diffraction intensities were recorded photographically (Ilford Industrial-G film, Eliot GX-6, 100μM focus, 39kV, 19mA, Franks focusing optics[7] at 4° C). At least two 1° oscillation photographs (exposure time 12 hours per degree) can be recorded from each crystal before X-ray damage is evident in the diffraction pattern. From 14 to 20° of oscillation data can be collected from some large crystals by translating the crystal to expose a fresh volume to X-rays every 24 hours. The crystals diffract to 3Å resolution with considerable loss in intensity with smaller spacings.

Data processing. Seventy-eight native and sixty-two derivative films were collected on an automated Supper oscillation camera and processed on an Optronics scanner (50μ raster).[8] Intensities from first and second films for each degree were merged by Fox and Holmes[9] scaling. Different films were scaled together by scaling the Wilson plots for the fully recorded intensities of each film to an average Wilson plot of the fully recorded intensities of the best measured films. Partially recorded intensities (≥50% recorded) were scaled to the fully recorded measurements by post-refinement methods.[10,11] Data processing statistics are shown in Table 1.

RESULTS

Difference Patterson. The location of six mercury atoms per asymmetric unit (two per monomer of the HA trimer) was determined from the difference Patterson at 3Å resolution. Inspection of the two Harker sections (Figure 1) at 5.5Å and 3Å illustrates the difficulty encountered in a Patterson of an 800,000 dalton asymmetric unit (200,000 dalton protein, 600,000 daltons solvent). Only two atoms (1 and 2 in Figure 2) have strong self Harker vectors. The four other solutions have at least one self Harker vector which is less significant than many other vectors on the Harker sections. Possible single

TABLE 1

DATA PROCESSING STATISTICS

Data set	No. films	Total reflections measured	Unique reflections	R_{sym}[2]	R_{whole}[3]	R_{asym}[4]
Native	78	408,672	83,015	.12	.11	.18
Derivative	62	338,738	78,077	.13	.13	.20

[1] $R = \Sigma|s_i I_{hi} - \bar{I}_h| / \Sigma \bar{I}_n$

s_i = scale factor for crystal i

I_{hi} = intensity of reflection h of crystal i

$\bar{I}_h$ = average intensity of reflection

[2] R_{sym} ---for symmetry related reflections on the same film

[3] R_{whole}--for fully recorded reflections from different films

[4] R_{asym} --for all reflections (≥50% recorded) on different films, final scaling

TABLE 2

HEAVY ATOM PARAMETERS USED IN PHASE DETERMINATION

atom site	x	y	z(Å)	occupancy[1]	B(Å^2)[2]
1	31.0	53.4	0.0[3]	0.47	22.7
2	25.1	54.0	21.8	0.61	19.9
3	0.4	42.9	-16.5	0.34	23.8
4	3.9	57.7	38.6	0.39	24.0
5	13.0	41.6	7.3	0.44	25.0
6	41.7	78.5	0.9	0.34	24.8

1 relative fractional occupancies (not absolute)
2 isotropic temperature factor
3 atom 1 z fixed at 0.0

Atoms 1, 2, 5 form one trimer and atoms 3, 4, 6 the second trimer. The Kraut R was 0.14, phasing power 1.36 for the 42,299 reflections used in the refinement with a figure of merit of 0.34.

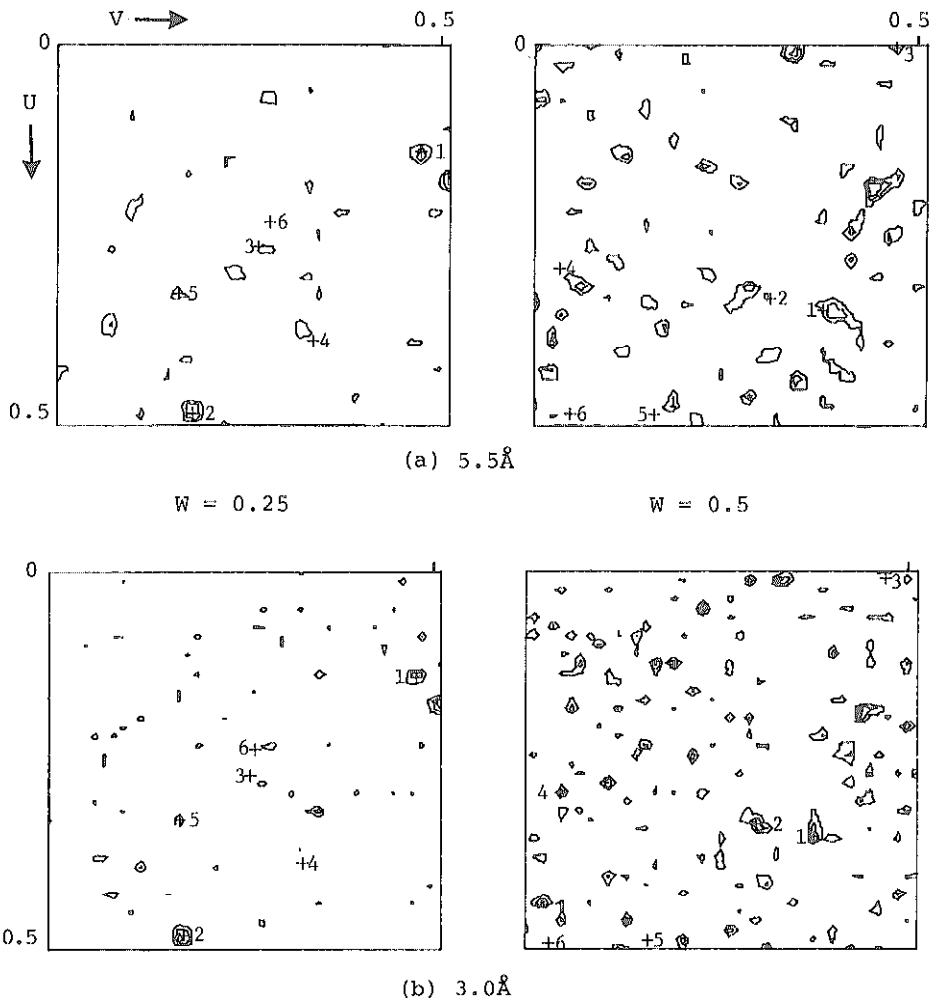

Fig. 1. Harker sections at 3Å and 5.5Å resolution of the difference Patterson. Self Harker vectors calculated from the refined atomic positions are indicated by a numbered cross on the Harker sections, W = 1/4 and W = 1/2.

atom solutions of the Patterson were determined by computing a symmetry minimum function yielding all potential x,y mercury atom coordinates consistent with the Harker sections at w = 1/4 and w = 1/2. Cross vectors for all pairs of such potential single atom solutions were then searched for in the three-dimensional Patterson. This search took the form of a symmetry minimum function in the Δz coordinate for an atom pair. From 32 potential single atom solutions, all cross vectors for a constellation of four atoms were found by this search. A fifth atom was found by symmetry minimum function searches in the Δz coordinates of the remaining large w = 0.5 Harker vectors which had no significant corresponding w = 0.25 vectors.

The remaining mercury position, vectors for which were not prominent on Harker sections, was found by considering the 583 largest peaks in the Patterson as potential cross vectors between the five known atoms and an unknown position. Simultaneously, a higher order image seeking function was written based on the double sort algorithm of Bricogne[1] by writing a new version of his routine "Generate". Given a subset of known atomic positions this function searched for cross vectors between a known atom and every position in the asymmetric unit (on a 2Å grid). The function determined that no solutions other than the original six atoms were present. In retrospect the function is capable of readily and unambiguously determining the location of all the heavy atoms from the coordinates of the first two found in the Harker search.

The final solution consists of two triangles of atoms 22.7Å and 57.3Å on an edge, the triangles separated by 13Å along the molecular three-fold axis. The stoichiometry of the phenyl glyoxal reaction is expected to be two phenyl glyoxals per arginine.[16] However, the separation of the two triangles seems to be too far apart for covalent attachment to the same arginine and may reflect two substituted arginines per HA_1 and HA_2 subunit.

Rotation function and the molecular three-fold symmetry axis. A single (sometimes split) prominent peak in the fast rotation functions[12] of the 3Å native, derivative and difference coefficients indicated a non-crystallographic three-fold symmetry axis nearly along the diagonal of the b,c plane (92° from the a-axis, 35.3° from the b-axis and 54.7° from the c-axis). The heavy-atom coordinates establish that this is a 120° non-crystallographic "packing" rotation axis relating two crystallographically identical molecules (such an axis will exist somewhere for any molecules related by a four-fold axis). The molecular three-fold axis, located from the heavy atom coordinates, is nearly

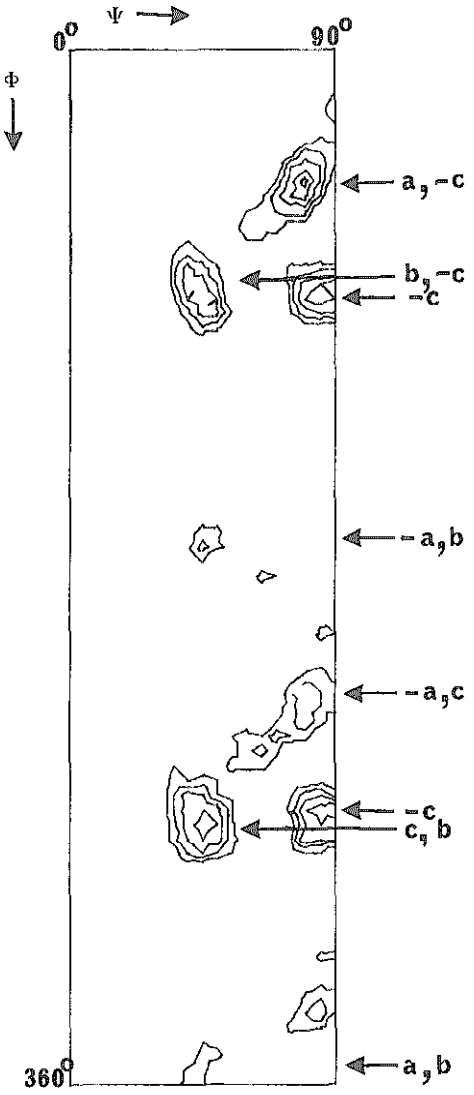

Figure 2. κ = 120° Rotation Function
A rotation function using the largest 1000 terms to 10Å resolution[13,14] shows the non-crystallographic molecular three-fold symmetry axis (located near the diagonal of the -â,b̂ plane) and non-crystallographic 120° "packing symmetry" axes. Polar angle definitions are given by Rossmann and Blow (1962).[13] Rotation functions calculated with radii from 15 to 100Å (R = 24Å shown) are similar. Planes (e.g., a, -c) or axes (c) are indicated on the rotation function.

along the diagonal of the a,b plane (53.4° from the a-axis, 141.6° from the b-axis and 79.9° from the c-axis).

A rotation function[13] based on the 1000 largest 10Å coefficients (Figure 2) shows both the packing and molecular non-crystallographic axes, the molecular axis being much less prominent than the packing axis (further analysis of the rotation function will be presented elsewhere).

Heavy atom parameter refinement and phase calculation. Final heavy atom parameters were determined by alternate cycles of least squares refinement and SIR phasing. All parameters refined well, including occupancy and isotropic temperature factors (Table 2).

The phases of 42,299 reflections used in the refinement (small measurements omitted) were calculated by the SIR method and had an overall figure of merit of 0.34.

Electron density map. An electron density map calculated with the SIR phases clearly indicated one molecule per asymmetric unit. The protein thus occupies only 22% of the total volume of the unit cell giving a small protein/unit-cell volume ratio well outside the normal for protein crystals.[15] The other possibility, that a second set of molecules does not fully occupy another location in the unit cell or is disordered, is not indicated by the Fourier maps.

The electron density map was averaged about the molecular three-fold[1] and the protein (carbohydrate) features were clearly enhanced above background solvent density by the averaging. The SIR map (without anomalous dispersion) is clearly subject to large errors as indicated by the solvent "noise". A comparison of two equivalent regions (4Å thick) of the averaged and unaveraged maps is shown in Figure 3. The connectivity has been increased and the background significantly decreased in the averaged map. On the unaveraged map the close contacts between the molecule and another related molecule can be seen.

DISCUSSION

Molecular shape. The molecule appears to be an elongate trimeric cylinder of approximately 130 Å length with radius varying from 15-38Å. Sections down the molecular three-fold axis often show the subunit boundaries with little density in the centre of the molecule (Figure 3). Other regions have tightly-packed subunits forming a triangular structure with a smaller radius (Figure 4).

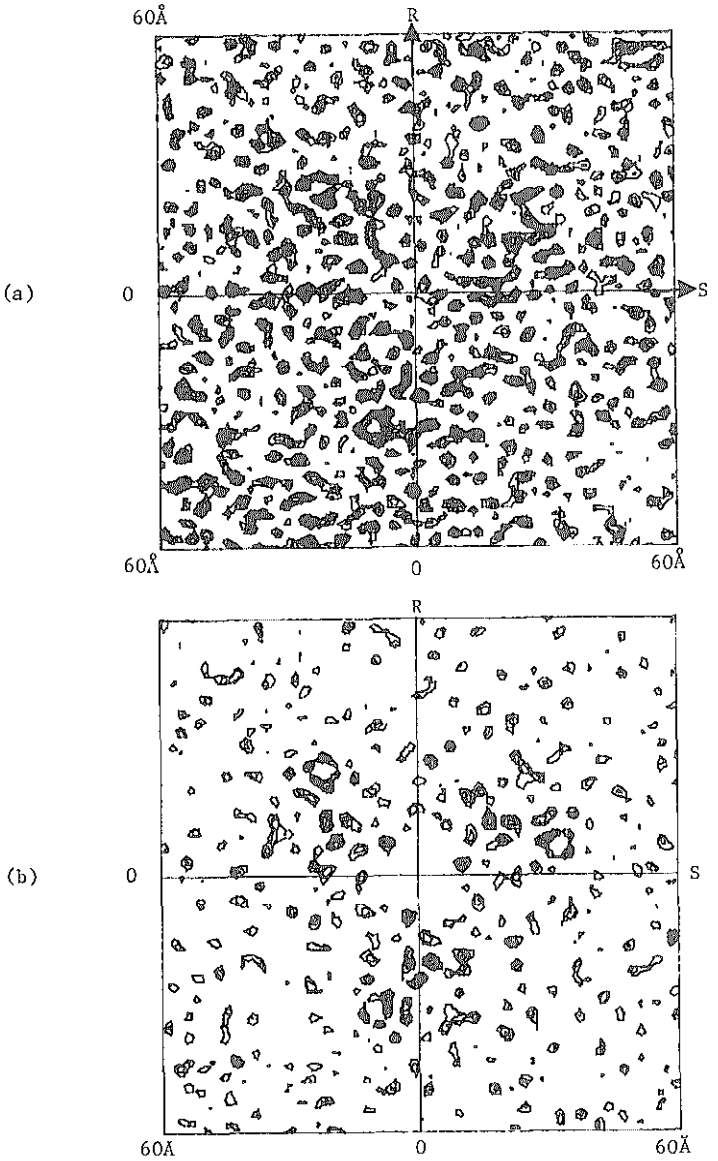

Fig. 3. Electron density maps of the unaveraged (a) and the three-fold averaged (b) SIR map at 3 Å resolution. The maps are viewed down the non-crystallographic molecular three-fold axis as defined in the text. The sections are 4Å thick and 120Å by 120Å centred on the three-fold. The large ring-like features viewed on the outside of each monomer represents artifact density around one triangle of mercury atoms (atoms 3, 4, 6, Table 2). R, S and T are the coordinate axes when viewed down the three-fold axis. The molecule extends from T = -45Å to +85Å. Sections above are T = -14 to -11Å.

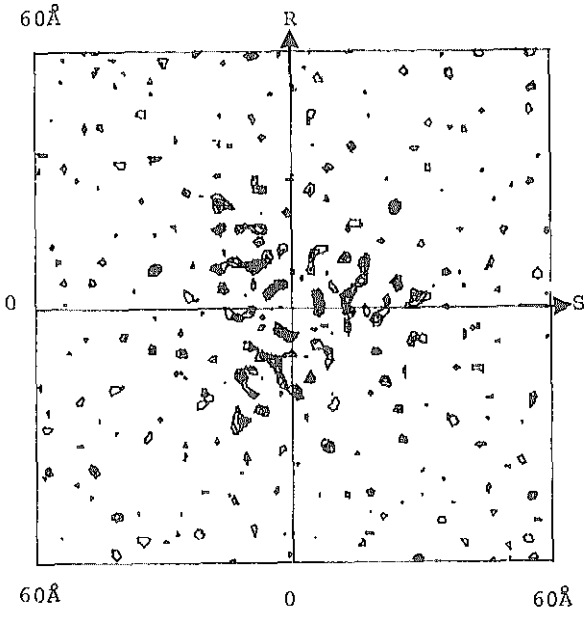

Fig. 4. Electron density of the three-fold averaged map at 3Å resolution in the "thin" end of the molecule where the subunits are more closely packed. The contour level is slightly higher than in Figure 3(b). A slice through the molecule 4Å thick down the non-crystallographic three-fold is shown for T = 14-17Å.

The average azimuthal position of the subunit varies slightly along the length of the three-fold axis. The molecule tends to be thicker at one end (70-80Å diameter), around the heavy atom locations (Figure 3), and thinner (30-50Å) at the other (Figure 4).

Detailed interpretation of the map has not been attempted pending phase refinement by non-crystallographic symmetry averaging.[1]

ACKNOWLEDGMENTS

We are indebted to many members of the Gibbs Structural Group for thoughtful discussions and to Edward Gordon and David Stevens for technical assistance.

REFERENCES

1. Bricogne, G. (1976) Acta Cryst. A32, 832-847.
2. Wiley, D.C. and Skehel, J.J. (1977) J. Mol. Biol. 112, 343-347.
3. Brand, C.M. and Skehel, J.J. (1972) Nature New Biol. 238, 145-147.
4. See this volume.
5. Kilbourne, E. (1969) Bull. Wld. Hlth. Org. 41, 643-645.
6. Monaco, H.L. (1978), Ph.D. thesis, Harvard University.
7. Harrison, S.C. (1968) J. Appl. Crystallogr. 1, 84-89.
8. Crawford, J.L. (1977) Ph.D. thesis, Harvard University.
9. Fox, G.C. and Holmes, K.C. (1966) Acta Cryst. 20, 886-891.
10. Schutt, C.E. (1976) PhD. thesis, Harvard University.
11. Winkler, F.K. et al., (1979) A35, 901-911.
12. Crowther, R.A. (1972) in The Molecular Replacement Method, ed. M.G. Rossmann, pp. 173-178, New York, Gordon and Breach.
13. Rossmann, M.G. and Blow, D.M. (1962) Acta Cryst. 15, 24-31.
14. Tollin, P. and Rossmann M.G. (1966) Acta Cryst. 21, 872-876.
15. Mathews, B.W. (1968) J. Mol. Biol. 33, 491-497.
16. Takahashi, K. (1977) J. Biochem. 81, 395-402.

DISCUSSION

Gething: What effect does the mercury phenyl glyoxyl substitution have on biological activity?

Wiley: We find no effect on hemagglutinating activity, but so far we have not done any infectivity assay.

Gibbs: I did not understand how the electron density map had polarity but was not 3-fold symmetrically superposed.

Wiley: The 3-fold axis of the molecule is not an axis of the crystal, so the operations of multiple crystals don't average. We can use this redundancy to get the phases.

Ward: The electron micrographs of Ian Griffith show the same bulging that your model has.

Wiley: The electron micrographs of Laver and Valatine of HA rosettes show a bulge at the outer end, which is the only clue at the moment as to which end is which.